Express Mail Label
No. EF230789724US

MIC-8

DATA COMPRESSION METHOD

Background of the Invention

[0001] This invention relates to a method for compressing data. More particularly, this invention relates to a method for reducing the coding length of data that is transformed into components, where the recipient is more sensitive to one component than the other. Most particularly, this invention relates to reducing the coding length of data that have been subjected to Fourier transformation.

[0002] Many types of analog data are digitized for transmission and processing. As is well known, the digitized representations of such data more accurately reflect the original analog signal as the number of bits per sample increases. One example of such an analog signal is speech, which, particularly if being digitized for a purpose involving the reconstitution of an analog signal for playback to human listeners, ideally should be represented sufficiently accurately to be understandable and at least relatively undistorted at the listener's end.

[0003] The number of bits per sample required for suitable reproduction of, e.g., speech, is high, and

runs up against bandwidth and other constraints. Therefore, ways are commonly sought to compress the digital data.

[0004] Moreover, a common and useful way of digitizing and transmitting an analog waveform, such as that representing speech or another physical phenomenon, is to subject the signal to Fourier transformation, such as by using a Fast Fourier Transform. The resulting transformed data are particularly well suited to processing and transmission. However, this actually compounds the compression problem, because M digital samples of the original analog waveform generate 2M transform coefficients (i.e., an M-sampled signal S(n) is transformed into 2M paired I/Q Fourier transform coefficients I(n) and Q(n)), doubling the coding length of the data.

[0005] It is apparent then, that it would be desirable to be able to reduce the coding length of Fourier transformed data.

Summary of the Invention

[0006] In accordance with the present invention, the coding length of M-sampled Fourier transformed data is reduced from 2M by as much as almost half by converting the Fourier transform coefficients into data representing magniture, or amplitude, of the original analog signal, and data representing the phase of the original analog signal.

[0007] The amplitude data preferably are transmitted at least substantially in their entirety. However, instead of transmitting the phase data in their

entirety, a smaller number of bits is used to transmit the phase data. This could be done by quantizing the phase to a smaller number of values than the amplitude. A more extreme compression could be obtained by transmitting only a single bit indicating the phase difference between the current sample and a related sample such as the previous sample. The single bit preferably would indicate whether the phase is advanced or retarded by a fixed amount as compared to the related sample. The fixed amount would be determined in advance and would be "known" to the receiving apparatus for use in reconstructing the original signal.

[0008] The invention works for, e.g., speech, because empirical observation shows that human listeners are relatively insensitive to the phase of a speech waveform. The invention may also work for music, although a discerning listener may detect imperfections. The invention may further work for non-sound waveforms, depending on what aspect of the waveform is most sensitive to coding precision.

[0009] Thus, in accordance with the invention, there is provided a method for compressing data for transmission to a recipient. The method includes transforming the data into at least two components, where the recipient is tolerant of variations in one of the components. A compressed representation of that one component is transmitted. Preferably, the compressed representation is data representing the change of that component from a related sample, such as the previous sample.

Brief Description of the Drawings

[0010] The above and other objects and advantages of the invention will be apparent upon consideration of the following detailed description, taken in conjunction with the accompanying drawings, in which like reference characters refer to like parts throughout, and in which:

[0011] FIG. 1 is a time-domain representation of a speech waveform;

[0012] FIG. 2 is a time-domain representation of a speech waveform created by digitizing the waveform of FIG. 1, quantizing it in the frequency domain using 2,000,001 possible phase values for each sample, and reconverting it to the time domain;

[0013] FIG. 3 is a time-domain representation of the difference (i.e., error) between the representation of FIG. 2 and the representation of FIG. 1;

[0014] FIG. 4 is a time-domain representation of a speech waveform created by digitizing the waveform of FIG. 1, quantizing it in the frequency domain using 15 possible phase values for each sample, and reconverting it to the time domain; and

[0015] FIG. 5 is a time-domain representation of the difference (i.e., error) between the representation of FIG. 4 and the representation of FIG. 1.

Detailed Description of the Invention

[0016] Empirical observation has shown that a human listener is relatively insensitive to phase errors during the playback of electronically processed speech signals. Therefore, in accordance with the present

invention, speech signals that have been processed electronically, particularly those that have been transformed into a format that actually increases the amount of data to be transmitted or played back, can be compressed with little perceivable loss in quality by reducing the amount of phase data that are transmitted or played back. Although the invention is described with respect to phase, similar compression might be achieved by reducing the amount of data representing any component with respect to which a recipient is tolerant of, or less sensitive to, variations. Moreover, while the invention is described with respect to speech, other audio data, and even other analog non-audio data such as seismic activity recordings, that can be resolved into components, to variations in one of which the recipient is relatively insensitive, can be compressed in accordance with the invention.

[0017] In a preferred embodiment of the invention, a speech waveform is digitized by an analog-to-digital converter, preferably with 16-bit accuracy, preferably at a sample rate of 8 kHz -- i.e., 8,000 16-bit samples preferably are collected each second, for a data rate in this preferred embodiment of 128,000 bits per second. These digitized speech data S(n) preferably are converted to the frequency domain through Fourier transformation, preferably using a Fast Fourier Transform. As a result, each 16-bit sample becomes two 16-bit Fourier transform coefficients I(n) and Q(n) -- i.e., there are 16,000 16-bit coefficients, for a data rate of 256,000 bits per second in this preferred embodiment.

[0018] The coefficients are then converted into magniture, or amplitude, R(n) and phase P(n), as follows:

[0019] $R(n) = ((I(n))^2 + (Q(n))^2)^{0.5}$

[0020] $P(n) = \tan^{-1}(I(n)/Q(n))$

[0021] The amplitude signal R(n) preferably is transmitted at least substantially in its entirety (i.e., at 128,000 bits per second in this embodiment). However, the phase signal P(n) preferably is compressed as described below.

[0022] Broadly considered, in accordance with the present invention, the phase signal P(n) is coarsely coded. For example, instead of transmitting sixteen bits per sample, only four bits per sample might be sent, and one method for deriving the four-bit values will be described below. Similarly, eight bits, or two bits, or any other number of bits fewer than sixteen bits could also be used to coarsely code the phase data. In the extreme as mentioned above, only one bit could be sent, indicating advance or retardation of the phase from a related sample, such as the previous sample. This method also will be discussed below.

[0023] In a first example, the spoken word "hello" was recorded as a .WAV file. The original waveform 10 is plotted in FIG. 1 as a function of the amplitude (in volts) versus time (as represented by the sample number). The .WAV file was then processed, using the MATLAB® Signal Processing Toolbox signal analysis utility available from The MathWorks, Inc., of Natick, Massachusetts, as follows:

[0024] First, the .WAV file was read into an array. Second, the time domain data in the array were

converted to the frequency domain, in rectangular or Cartesian coordinates, using a Fast Fourier Transform. Next, the Cartesian frequency domain data were converted to polar coordinates, where the radius represented the magniture or amplitude, and the angle, ranging from $-\pi$ to $+\pi$, represented the phase. The amplitude was transmitted with full precision.

[0025] Each phase sample was then quantized to one of a plurality of discrete values by selecting an integer N, normalizing the value of the phase sample to between -1 and +1 by dividing it by π, multiplying the normalized phase value by N, rounding the product to the nearest integer, dividing the rounded product by N and finally multiplying by π.

[0026] It will be appreciated that the rounded product of N and the normalized phase is an integer between -N and +N, which can have 2N+1 possible values (-N, ... , -2, -1, 0, 1, 2, ... , N). Dividing each of that many possible values by N and multiplying by π will not change the number of possible values. Therefore, the final result is that each phase sample is quantized to one of 2N+1 values. It will further be appreciated that the accuracy of the representation of the phase data by the quantization values increases as N increases.

[0027] Quantization was tried with N=1,000,000 (2,000,001 possible quantization values) and N=7 (15 possible quantization values). In each case the result, along with the full-precision amplitude data, was converted back to the time domain using an inverse

Fast Fourier Transform, to produce a .WAV file that could be played back.

[0028] The resulting waveform 20 for the case where N=1,000,000 is plotted in FIG. 2 as a function of amplitude (in volts) versus time (as represented by the sample number). Visual comparison reveals that waveform 20 of FIG. 2 is virtually indistinguishable from original waveform 10 of FIG. 1. Empirically, it was observed upon playing back of the two .WAV files that to a human listener they were aurally indistinguishable as well. Indeed, the error 30 between waveform 20 and waveform 10, obtained by subtraction, is shown in FIG. 3, and has a maximum value of 8×10^{-7} volts.

[0029] The resulting waveform 40 for the case where N=7 is plotted in FIG. 4 as a function of amplitude (in volts) versus time (as represented by the sample number). Visual comparison reveals that waveform 40 of FIG. 4 is similar to original waveform 10 of FIG. 1, but not so indistinguishable from waveform 10 as, e.g., waveform 20 was. Indeed, the error 50 between waveform 40 and waveform 10, obtained by subtraction, is shown in FIG. 5, and has a maximum value of close to 0.1 volts, or about 10% of the original signal. Nevertheless, it was observed empirically upon playing back of the resulting .WAV file that it sounded to a human listener virtually identical to the .WAV file represented by waveform 10.

[0030] Significantly, storage or transmission of the full precision Fourier-transformed signal typically would require 32 bits (16 bits for each of I(n), Q(n) or R(n), P(n) signal pairs). On the other hand,

storage or transmission of waveform 40, which empirically sounds the same, would require only 20 bits (16 bits for R(n) and 4 bits for (P(n)).

[0031] In a second example, the spoken word "hello" again is recorded as a .WAV file (FIG. 1). The .WAV file is then processed, using the MATLAB® Signal Processing Toolbox signal analysis utility, as follows:

[0032] First, the .WAV file is read into an array as before. Second, as before, the time domain data in the array are converted to the frequency domain, in rectangular or Cartesian coordinates, using a Fast Fourier Transform. Next, the Cartesian frequency domain data are converted to polar coordinates, where, as above, the radius represents amplitude, and the angle represents phase. The amplitude is transmitted with full precision.

[0033] With respect to the phase, the value of the first (reference) sample preferably is set to zero. Thereafter, for each subsequent sample, a single bit preferably is transmitted, indicating whether the phase is advanced or retarded by some preferably fixed amount as compared to a related sample, which could be the previous sample, the next sample or another subsequent sample, the same sample in a previous or subsequent block of speech, or a sample related in some other predetermined way to the current sample. For example, a "1" could indicate that the phase is advanced while a "0" could indicate that the phase is retarded, or vice-versa. In a case where there is no change in the phase over several samples, the phase bits alternate between "1" and "0", alternately advancing and retarding the

phase by the same amount, so that on average there is no phase change.

[0034] The value of the "fixed amount" of phase change is determined empirically and "made known" in advance to the receiving/playback apparatus. The value must be small enough to produce acceptable fidelity (i.e., the value cannot be so large that the system does not register phase changes), but large enough to allow the system to respond (i.e., given that the value is fixed, the value cannot be so small that when a change is registered, the output change is insufficient to approximate the real change).

[0035] On the one hand, there is the question of how much of a phase change there has to be before the system reacts. On the other hand, if the system is to react, and is going to react by a fixed amount, then that fixed amount has to be some substantial portion of the full excursion of the phase data between the maximum and minimum phase values for the entire waveform. This requires knowing the likely maximum difference between phase samples. Depending on the system design, it may be that there is some known correlation between frequency samples. If so, it may be possible to select the same frequency sample from successive blocks of speech and encode only the difference in phase between them. Thus the invention likely would not work well for signals where there is little or no correlation between samples and the phase could assume any value from one sample to the next.

[0036] Another possibility may be to accumulate or "batch up" phase changes without transmitting them, either for a predetermined number of samples (e.g.,

covering 20 ms of speech data), or until the predetermined fixed amount is reached, and then to transmit the one or a few bits indicating that there is an increase or decrease of that amount (or no change if after a predetermined number of samples there is no net change).

[0037] If necessary, more than one bit could be used, to indicate by how many of the fixed increments the phase has changed. If one bit is used, the entire signal could be transmitted in this example using 17 bits instead of 32 bits, for a reduction by almost half of the full coding length. Generally speaking, the maximum expected difference between two phase values must be encodable by the largest value of the phase sample signal (which is a function of the number of bits used and the value of the increment the multiple of which they represent).

[0038] Any other compression scheme that takes advantage of listeners' relative insensitivity to phase variations in speech, or possibly other types of audio waveforms such as music, can be used. Similarly, if waveform data or any other type of data, such as seismic activity recordings, can be broken down into two or more components, where the recipient of the data is relatively tolerant of, or insensitive to, variations in one of those components, then in accordance with the invention, the data can be compressed by more coarsely coding that component to variations of which there is less sensitivity.

[0039] It should be noted that although the discussion above indicates that the amplitude data, or data representing any component to variations in which

a recipient would be sensitive, is transmitted with full precision, or with at least substantially full precision, that is not meant to exclude the possibility that any data compressed by the method according to this invention might be further compressed by one of the well known general compression schemes commonly in use, such as MP3. Thus, in the speech examples set forth above, the output of the method according to this invention would be a full-precision (or substantially full-precision) amplitude signal and a compressed phase signal. That output could subsequently be subjected to one of the aforementioned general compression schemes as well.

[0040] At the receiving end, a signal compressed according to the present invention would be simply played back if compressed according to the first example, or, if compressed according to the second example, subject to reconstruction by advancing or retarding the phase for each sample as indicated by the compressed data, and then played back. If one of the aforementioned general compression schemes is used on the output of the method of this invention, then at the receiving end, the corresponding decompression scheme would be used first, and then the signal output by the present invention would be played back as just described.

[0041] Thus it is seen that the coding length of digitized data, particularly Fourier-transformed data, and particularly such data representing speech, can be decreased by up to almost half in accordance with the present invention. One skilled in the art will appreciate that the present invention can be practiced

by other than the described embodiments, which are presented for purposes of illustration and not of limitation, and the present invention is limited only by the claims which follow.